

*Citation for published version:*

Ball, A 2013, 'Research Data Management: Policy, Planning and Practice', ASLIB Engineering and Technology Afternoon Seminar: Research Data Management and Patent Searching, Guildford, UK United Kingdom, 22/05/13.

*Publication date:*

2013

*Document Version*

Early version, also known as pre-print

[Link to publication](#)

*Publisher Rights*

CC BY

**University of Bath**

**Alternative formats**

If you require this document in an alternative format, please contact:  
[openaccess@bath.ac.uk](mailto:openaccess@bath.ac.uk)

**General rights**

Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

**Take down policy**

If you believe that this document breaches copyright please contact us providing details, and we will remove access to the work immediately and investigate your claim.

# Research Data Management

Policy, Planning and Practice

Alex Ball

22 May 2013

## Abstract

Research data management has always been important, but the need to support it with specialist facilities and formal processes has only been widely recognised since the 1990s. Progress has been accelerating since then, with data management planning becoming a common feature of the funding application process in the late 2000s and institutional research data policies appearing from 2011. This talk considers the nature of research data, why it should be managed and shared, how the research community is responding in policy, planning and practice, and how academic libraries can provide support.

This talk was given to the ASLIB Engineering and Technology group at the University of Surrey on 22 May 2013.

## Contents

<b>1</b>	<b>Research Data</b>	<b>2</b>
<b>2</b>	<b>Management</b>	<b>5</b>
<b>3</b>	<b>Policy</b>	<b>7</b>
<b>4</b>	<b>Planning</b>	<b>9</b>
<b>5</b>	<b>Practice</b>	<b>11</b>

**Acknowledgement.** Some of the content of this talk derives from 'Research Data Management for Librarians' by Sarah Jones, Marieke Guy and Miggie Pickton: <http://www.dcc.ac.uk/training/rdm-librarians>

# 1 Research Data

‘Research data’ can be a rather nebulous term so it’s worth spending time working out what we mean by it.¶

According to the MANTRA project . . . <sup>1</sup>

- Research data are collected, observed or created, for the purposes of analysis to produce and validate original research results.¶

According to the *Scientific Data Application Profile Scoping Study* . . . <sup>2</sup>

- The evidence base on which academic researchers build their analytic or other work, where this evidence base is typically gathered, collated and structured according to declared and accepted protocols.¶

According to EPSRC . . . <sup>3</sup>

- Recorded, factual material commonly retained by and accepted in the [re-search] community as necessary to validate research findings.¶

So many things can be research data:

- Simulation data, models and software
- Survey results and interview transcripts
- Instrument measurements
- Machining/measurement parameter data
- Still images, video and audio
- Experimental observations and field notes
- Text documents, spreadsheets, databases
- System log files
- Specimens, samples, slides, artefacts
- Sketches, diaries, lab notebooks . . .

*...and that's just in Engineering!*

Research data is not a type of *thing*, it is a type of *use*.

I was involved in the JISC-funded ERIM Project which looked at engineering research data management, and when we tried to understand the data coming out of engineering research, we found quite a few different roles data could play:¶

Data Object (physical or digital)

---

<sup>1</sup><http://data.lib.edina.ac.uk/mantra/researchdataexplained.html>

<sup>2</sup><http://www.ukoln.ac.uk/projects/sdapss/>

<sup>3</sup><http://bit.ly/130P5Kv>

- **Data Record**
  - **Research Object Data Record** – *a resource which is the thing being studied*
  - **Experimental Apparatus Data Record** – *a resource used to measure, mine or analyse the object of study*
  - **Research Data Record** – *the output of an experiment, observation, simulation or processing operation*
  - **Context Data Record** – *a resource intended to explain/justify the research activity*
    - \* **Associative Data Record** – *a resource explaining the relationships between other records*
    - \* ...
- **Context Data Object** – *a resource shedding light on the research activity*
  - **Unintentional Context Data Object** – *a resource created ‘accidentally’ as a by-product of the research activity*
  - **Intentional Context Data Object** – *a resource created deliberately to describe the research activity (e.g. Context Data Record)*

All of which shows you have to think broadly about research data and the role it has in making research intelligently open. I’m borrowing that phrase from a report published by the Royal Society last year ¶ *Science as an Open Enterprise*.<sup>4</sup>

Intelligently open data is

- **Accessible:** easy to find
- **Intelligible:** others should be able to understand it
- **Assessable:** both on its own merits and in context of authors’ interests
- **Usable:** in a sensible format, properly documented

The report contains several strong statements about the importance of sharing research data:

- Public communication of scientific knowledge should not simply disclose conclusions but also communicate the reasoning and evidence that underlie them. – page 38



There is an implication in the report that disseminating conclusions without opening up the evidence for inspection amounts to scientific malpractice, and there have certainly been high profile cases recently illustrating the dangers of not sharing data. ¶

Reinhart and Rogoff (2010) said public debt at 90% GDP was tipping point for recession (Figure 1).<sup>5</sup> This became a core justification for austerity measures here and overseas. BUT

<sup>4</sup><http://royalsociety.org/policy/projects/science-public-enterprise/report/>

<sup>5</sup>Reinhart, C. & Rogoff, K. (2010). *Growth in a Time of Debt* (Working Paper 15639). Cambridge, MA: National Bureau of Economic Research. Available from <http://www.nber.org/papers/w15639.pdf>

	B	C	I	J	K	L	M
2			Real GDP growth				
3			Debt/GDP				
4	Country	Coverage	30 or less	30 to 60	60 to 90	90 or above	30 or less
26			3.7	3.0	3.5	1.7	5.5
27	Minimum		1.6	0.3	1.3	-1.8	0.8
28	Maximum		5.4	4.9	10.2	3.6	13.3
29							
30	US	1946-2009	n.a.	3.4	3.3	-2.0	n.a.
31	UK	1946-2009	n.a.	2.4	2.5	2.4	n.a.
32	Sweden	1946-2009	3.6	2.9	2.7	n.a.	6.3
33	Spain	1946-2009	1.5	3.4	4.2	n.a.	9.9
34	Portugal	1952-2009	4.8	2.5	0.3	n.a.	7.9
35	New Zealand	1948-2009	2.5	2.9	3.9	-7.9	2.6
36	Netherlands	1956-2009	4.1	2.7	1.1	n.a.	6.4
37	Norway	1947-2009	3.4	5.1	n.a.	n.a.	5.4
38	Japan	1946-2009	7.0	4.0	1.0	0.7	7.0
39	Italy	1951-2009	5.4	2.1	1.8	1.0	5.6
40	Ireland	1948-2009	4.4	4.5	4.0	2.4	2.9
41	Greece	1970-2009	4.0	0.3	2.7	2.9	13.3
42	Germany	1946-2009	3.9	0.9	n.a.	n.a.	3.2
43	France	1949-2009	4.9	2.7	3.0	n.a.	5.2
44	Finland	1946-2009	3.8	2.4	5.5	n.a.	7.0
45	Denmark	1950-2009	3.5	1.7	2.4	n.a.	5.6
46	Canada	1951-2009	1.9	3.6	4.1	n.a.	2.2
47	Belgium	1947-2009	n.a.	4.2	3.1	2.6	n.a.
48	Austria	1948-2009	5.2	3.3	-3.8	n.a.	5.7
49	Australia	1951-2009	3.2	4.9	4.0	n.a.	5.9
50							
51			4.1	2.8	2.8	=AVERAGE(I:30;L:44)	

Figure 1: Coding error in the Reinhart–Rogoff spreadsheet

- contrary data from post-war Canada, Australia and New Zealand excluded
- contrary data given less weight
- coding error

Initially no-one else could reproduce the results. The mistakes and odd choices only came to light once the data was made public. (And if you correct for them, you find growth of 2.2% at this level of debt and no tipping point at all.)<sup>6</sup>

It gets worse.¶

- 2002: Jan Hendrick Schön found to have falsifying/fabricating results in *at least 17 revolutionary* papers in solid state physics.
- 2006: Woo Suk Hwang found to have faked data relating to *groundbreaking* cloning and stem-cell research.
- 2011: Diederik Stapel found to have fabricated data underlying 30 peer-reviewed papers in clinical psychology.

This last example has been particularly influential in shifting attitudes... implicit trust broken... now culture of transparency developing. And that is a nice example in microcosm of how attitudes are shifting across the board. Nowhere is that shift more apparent than they way research funders are becoming increasingly concerned with research data management.

<sup>6</sup>Konczal, M. (2013). Researchers finally replicated Reinhart-Rogoff, and there are serious problems. *Next New Deal* [blog]. Available from the Roosevelt Institute website: <http://www.nextnewdeal.net/rortybomb/researchers-finally-replicated-reinhart-rogoff-and-there-are-serious-problems>

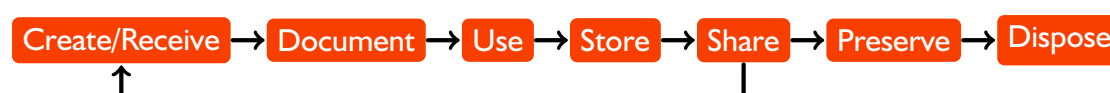


Figure 2: A research data management lifecycle

## 2 Research Data Management

- **mid-1990s:** NERC requires sharing of data funded by its grants
- **2000:** ESRC requires sharing of data funded by its grants
- **2004:** OECD *Declaration on Access to Data from Public Funding* – UK signs up; UNESCO *Policy Guidelines for the Development and Promotion of Governmental Public Domain Information* – includes research data
- **2005–6:** RCUK issues a position statement *that talks about research outputs, but focuses on Green Open Access publishing*
- **2007:** BBSRC, MRC, Wellcome require sharing of data funded by its grants
- **2008:** RIN *Stewardship of Digital Research Data*; AHRC requires sharing of data funded by its grants
- **2011:** STFC requires sharing of data funded by its grants; EPSRC requires institutions to perform RDM

This timeline shows some major milestones...simplified...starts bottom-up in disciplines with unreproducible data, then top-down as a political priority. Also, starts with burden for RDM on funder, with co-operation from researchers, to researchers having to more for themselves, and ends up with EPSRC putting the burden squarely on institutions.

So what does research data management involve?¶

- the active management and appraisal of data over the lifecycle of scholarly and scientific interest. • – Digital Curation Centre

Figure 2 gives an idea of the stages involved...

¶ At the *data collection* stage, activities include

- Gaining consent *from individual participants*, contractual permissions *from organisations*, licences *to use existing data*
- Collecting data *according to well considered and justified procedures*
- Performing quality assurance and control
- Recording data in useful formats– *implications about software used*

¶ At the *documentation* stage, one is Recording all information needed to understand and use the data, e.g. codes used to categorise survey responses, special numeric values with non-numeric semantics (such as 'error', 'dummy value').

- at the time – it's harder to do later
- according to metadata standards – *see for example <http://www.dcc.ac.uk/resources/metadata-standards>*
- including local file naming, version control or structural conventions *which are all Good Things*

¶ At the point of *use*, activities include

- Making analysis and processing reproducible: appropriate software, recorded steps (*e.g. workflow automation systems such as Taverna*)
- Using secure, robust collaboration systems – *security considerations will influence choices of where data is stored, which devices are used for access, who has which permissions*
- Version control: synchronisation, single working copy

¶ When looking at how data is *stored*, one should be

- Using managed storage rather than removable media
- Backing up data so there are
  - at least 3 copies of a file
  - on at least 2 different media
  - with 1 copy offsite

¶ When it comes to *sharing* and *preserving* data, the researcher's job should be easy if they have managed their data well up to this point. The main task will be **Deciding** what data should be published, what data should be in a dark archive, and what should be **discarded** – *one can't keep everything because it would cost too much and make it hard to find relevant data*

- What must be kept (legally, contractually, for scientific integrity)?
- What must be destroyed (legally, contractually)?
- What is of particular scientific, historical or economic value/interest?
- What is unique, or cannot be reproduced?
- Is there enough documentation to make keeping the data worthwhile?

More advice is available from the DCC.<sup>7</sup>

¶ While sharing and preservation are distinct operations, they are usually both accomplished by depositing data in a data archive or institutional data repository. This involves tasks such as

---

<sup>7</sup>How to Appraise and Select Research Data for Curation': <http://www.dcc.ac.uk/resources/how-guides/appraise-select-data>

- **Selecting an appropriate repository:** Databib, Re3data – *actually it is a good idea to do this very early on, as some have experts that can help researchers with the whole data management lifecycle*
- **Packaging data in an appropriate submission format** *e.g. XFDU, DDI, METS, SWORD*
- **Obtaining a persistent ID (DOI, Handle) for the data**
- **Linking data to papers it supports**

It is important that all these tasks are performed well, but this means co-ordination across many actors... Institutions and departments need to set up policies and procedures, and provide infrastructure... Research managers and principal investigators need to plan how their projects will be run in advance. And in the course of projects, tools are needed to make it simple and easy to follow the plan.

### 3 Research Data Management Policy

Over the past two or three years, institutions have started adopting research data policies. Such policies are there to ensure compliance with legal and funder requirements. They demonstrate the institution takes these matters seriously and provide a mandate for activity.

¶ The main funder for Engineering research is of course the EPSRC... uniquely, it places the burden of responsibility on institutions. Here are its expectations, paraphrased for brevity:

1. *Research organisations (ROs) to raise awareness of data sharing responsibilities and issues.*
2. *Publications should link to underlying data.*
3. *ROs must keep track of their research datasets and requests for them.*
4. *Born-analogue data must also be shareable on request.*
5. *ROs must provide open, online catalogues of their data; digital data must be given a robust ID.*
6. *Access restrictions should be clear and justified.*
7. *ROs must provide access to data for 10 years from last access.*
8. *ROs must curate their research data.*
9. *ROs must pay for this from their existing public funding streams.*

This gives you an idea of the kinds of things a research data policy should cover, but what should it look like? Here are a few contrasting examples.<sup>8</sup>

¶ The University of Oxford got in early with a short statement of commitment, leaving the full policy until it had sufficient infrastructure in place:

---

<sup>8</sup>The DCC maintains a list of institutional data policies: <http://www.dcc.ac.uk/resources/policy-and-legal/institutional-data-policies/uk-institutional-data-policies>



- Statement of Commitment (2009)<sup>9</sup>

“The University of Oxford is committed to supporting researchers in appropriate curation and preservation of their research data, and where applicable in accordance with the research funders’ requirements. It recognises that this must be achieved through the deployment of a federated institutional data repository. . . .”

- Policy on the Management of Research Data and Records (2012):<sup>10</sup> 12 points, detailing with expectations for research data, retention periods, roles and responsibilities, etc.

¶ The University of Edinburgh took a different approach. It adopted an aspirational policy, recognising that it will be several years before the infrastructure is in place to comply with it. It is a simple ten-point policy; I’ve only show the first three on the slide.<sup>11</sup>

1. Research data will be managed to the highest standards throughout the research data lifecycle as part of the University’s commitment to research excellence.
2. Responsibility for research data management through a sound research data management plan during any research project or programme lies primarily with Principal Investigators (PIs).
3. All new research proposals [from date of adoption] must include research data management plans or protocols that explicitly address data capture, management, integrity, confidentiality, retention, sharing and publication.
4. . . .

¶ The University of Hertfordshire already had a policy for its administrative data. It extended it to cover research data as well; on the slide is most of the section specifically on research data (I’ve added my own emphasis):<sup>12</sup>

Data management is an *essential and integral part* of the responsible conduct of research. The University is responsible for:

1. ensuring effective data management to meet internal and external requirements, including enabling the re-use of research data and freely available public access to research data outputs in accordance with *national and funding body policies*;
2. retention of research data in sufficient detail for a defined period to enable appropriate *responses to any questions* about accuracy, authenticity, primacy and compliance with legal and regulatory requirements governing the conduct of research;

<sup>9</sup>Oxford statement: <http://www.ict.ox.ac.uk/odit/projects/datamanagement/>

<sup>10</sup>Oxford data policy: [http://www.admin.ox.ac.uk/media/global/wwwadminoxacuk/localsites/researchdatamanagement/documents/Policy\\_on\\_the\\_Management\\_of\\_Research\\_Data\\_and\\_Records.pdf](http://www.admin.ox.ac.uk/media/global/wwwadminoxacuk/localsites/researchdatamanagement/documents/Policy_on_the_Management_of_Research_Data_and_Records.pdf)

<sup>11</sup>Edinburgh data policy: <http://www.ed.ac.uk/schools-departments/information-services/about/policies-and-regulations/research-data-policy>

<sup>12</sup>Hertfordshire data policy: <http://sitem.herts.ac.uk/secreg/upr/IM12.htm>

**Table 1: CARDIO statements**

Organisation	Technology	Resources
<ol style="list-style-type: none"> <li>1. Data Ownership and Management</li> <li>2. Data Policies and Procedures</li> <li>3. Data Policy Review</li> <li>4. Sharing of Research Data/Access to Research Data</li> <li>5. Preservation and Continuity of Research</li> <li>6. Internal Audit of Research Activities</li> <li>7. Monitoring and Feedback of Publication</li> <li>8. Metadata Management</li> <li>9. Legal Compliance</li> <li>10. Intellectual Property Rights and Rights Management</li> <li>11. Disaster Planning and Continuity of Research</li> </ol>	<ol style="list-style-type: none"> <li>1. Technological Infrastructure</li> <li>2. Appropriate Technologies</li> <li>3. Ensuring Availability</li> <li>4. Managing data integrity</li> <li>5. Obsolescence</li> <li>6. Managing technological change</li> <li>7. Security Provisions</li> <li>8. Security Processes</li> <li>9. Metadata tools</li> <li>10. Institutional Repository</li> </ol>	<ol style="list-style-type: none"> <li>1. Data Management Costs and Sustainability</li> <li>2. Business Planning</li> <li>3. Technological Resources Allocation</li> <li>4. Risk Management</li> <li>5. Transparency of Resource Allocation</li> <li>6. Sustainability of Funding for Data Management and Preservation</li> <li>7. Data Management Skills</li> <li>8. Number of Staff for Data Management</li> <li>9. Staff Development Opportunities</li> </ol>

<http://cardio.dcc.ac.uk/>

### 3. for supporting investigation into any allegations of misconduct or regulatory breach

It also wrote a guidance document explaining what research data management is and how to write a data management plan.

It's one thing to write these policies and another to put in place the infrastructure allowing researchers to follow it. Fortunately there are tools to help.

¶ To get a handle on what data an institution holds, there's the *Data Asset Framework*, or DAF. This was originally intended as a methodology for assembling an inventory of data assets, based on desk research, interviews and questionnaires. But what its users found really valuable were the insights it gave them into the state of current practice and the scale and variety of the data assets out there in the wild.

¶ Taking things one step further there's *CARDIO*, a sort of health check for institutional RDM. It comes in two flavours. The first is very quick and easy: ten multiple-choice questions that guide you through the main areas of data management and invite you to reflect on how well you're doing in each of them. ¶ The second is rather more thorough. It invites you and other stakeholders to assess the institution's performance in 30 different areas (Table 1), and provides facilities for getting a consensus view and formulating a concrete action plan. This might include writing policies on IPR or risk management, rethinking how IT facilities are financed, providing new infrastructure such as a data catalogue or repository, or providing data management training.

Co-incidentally, these are all things that principal investigators need to know about when they are writing funding proposals, because they come up as part of the data management plan.

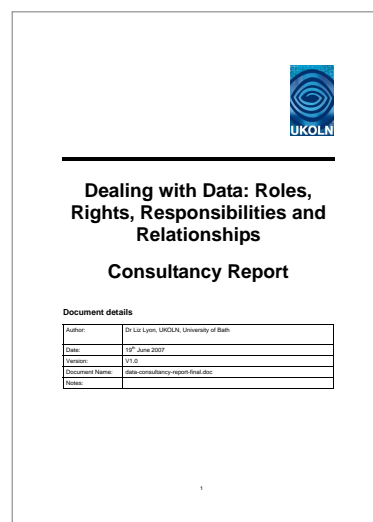
## 4 Research Data Management Planning

¶ Data management plans started showing up as a required part of funding proposals in the mid-2000s. Liz Lyon saw this when writing her *Dealing with Data* report and thought they were a jolly good idea.

- 6 **Recommendation 9.** Each funded research project, should submit a structured Data Management Plan for peer-review as an integral part of the application for funding. — Liz Lyon (2007), *Dealing with Data: Roles, Rights, Responsibilities and Relationships* (University of Bath)

Why? ¶ Writing and using a Data Management Plan helps

- to co-ordinate the actions of data stakeholders
- to ensure all necessary tasks are accomplished
- to ensure data are properly curated
- with releasing data in a timely fashion
- with sharing data as openly as possible – *Engineering could learn from medical research in how to share sensitive information*
- with preserving data for future use



Data management plans have a lifecycle too... The first stage is at the point of bidding for funding... These plans demonstrate that the applicant has thought about data issues, so they won't waste time collecting data that already exist, and the new data they produce will be usable and shareable. Once the bid has been accepted, the plan needs to be firmed up to reflect the practical realities of the research. Currently only NERC mandates this stage but it needs doing. It's also a good idea to review it periodically throughout the project to make sure that it is being followed, and make any necessary adjustments.

Towards the end of the project, the DMP becomes a useful part of the data management record, which can be handed over to a data centre or repository as evidence for the provenance of the data. The data centres and repositories will themselves have data management plans that are mostly focused on curation and long-term preservation.

¶ The DCC runs a service called DMPonline, which allows people to:

1. create, store and update Data Management Plans
2. meet both institutional and funders' data-related requirements, *all in one go*
3. receive specific guidance from funders and institutions
4. export Data Management Plans in various formats

¶ Here is what it looks like (Figure 3)...

The EPSRC does not currently require researchers to complete a DMP as part of the bidding process, but institutions might. Chances are that you will not be asked by engineers about DMPs too often, but when they do ask you they will need a lot of help. ¶

Various types of support could be provided by libraries:

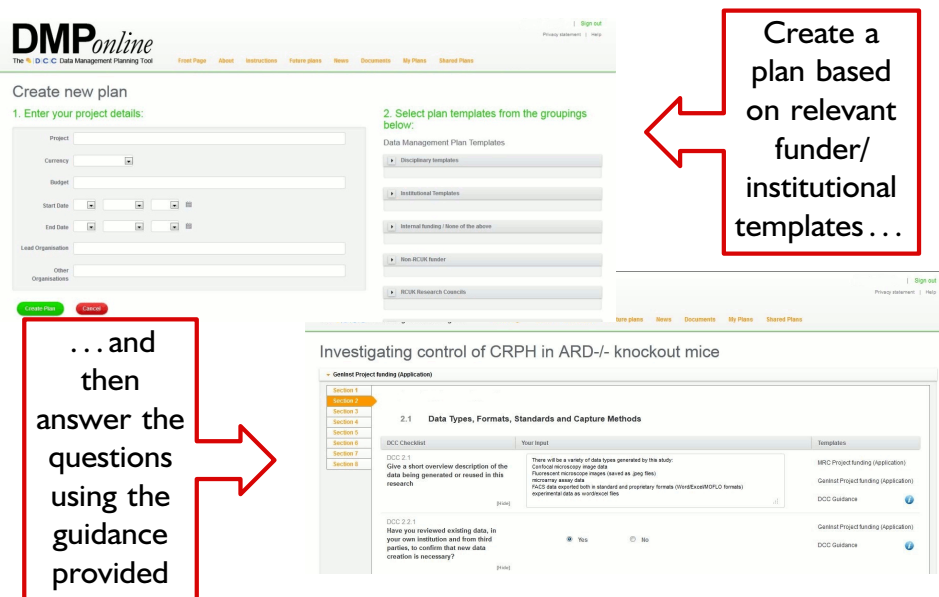


Figure 3: How DMPonline works

- Guidelines and templates on what to include in plans
- Example answers, guidance and links to local support
- A library of exemplar DMPs
- Training courses and guidance websites<sup>13</sup>
- Tailored consultancy services
- Online tools (e.g. customised DMPonline)

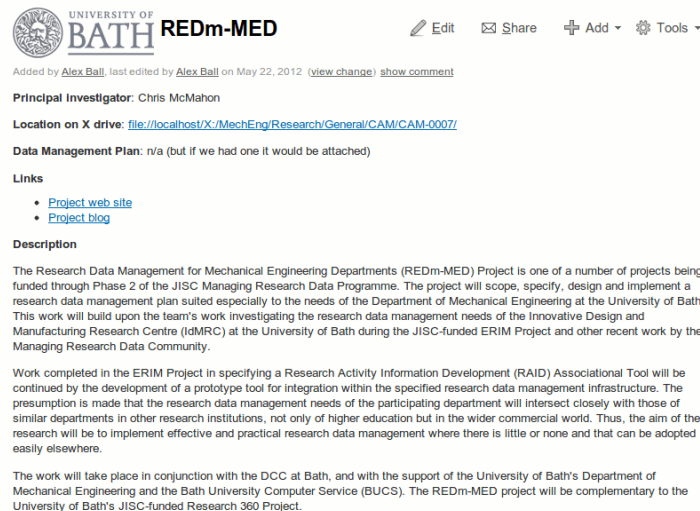
Why would institutions make researchers write DMPs even if the funder doesn't ask for them? They are useful because they reveal researchers' real data management needs; they're very useful for planning storage provision, for example.

## 5 Research Data Management in Practice

Ideally data management plans for funded projects should be open for examination; and not just DMPs but also:

- Project proposal (pre-award)
- Detailed project plan (post-award)
- Project record manifest – *I'll explain this in a moment*
- Confidentiality agreements *which may themselves be confidential!*
- IPR statements, licences – *strongly affect how data may be reused*

<sup>13</sup>The DCC maintains a list of institutional RDM guidance websites: <http://www.dcc.ac.uk/resources/policy-and-legal/rdm-guidance-webpages/rdm-guidance-webpages>



**Figure 4:** Departmental Research Data Management Wiki

But that's not always possible, especially in Engineering. Some projects are so sensitive that even the data management plan might give away too much information ...

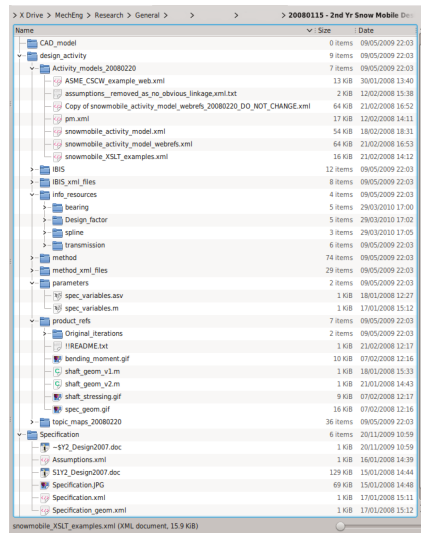
¶ For the Department of Mechanical Engineering in Bath, we decided we should have a public space for listing projects (Figure 4), but if DMPs were sensitive, they should be kept in a secure space, and the public page would simply point to that location and explain who was allowed to access it. We also encouraged researchers to provide a redacted version of sensitive documents on the public wiki wherever possible. All this was done in the context of a project called REDm-MED, which looked at the practical business of research data management.

As I mentioned at the start, Engineering data is incredibly diverse. Looking at the work of just one centre within the Department of Mechanic Engineering, it seemed every project was working with a different set of formats and using a different workflow. It's unlikely that anyone coming to directory of data like this (¶ Figure 5), even the researchers themselves a few months on, would know what it all means and how it fits together.

So, in ERIM and REDm-MED we decided the way to solve this would be to create a Project Record Manifest (Figure 6).

The main component of this (let me zoom in §) is a table listing all the records associated with a project, showing the record title, file name and location, owner and contact details, record type, and confidentiality status. With this at least we have a chance of working out what's what. But filling out that table is laborious, and very easy to forget to do until the task becomes monumental. Plus, there's a lot a mere table can't do, such as indicate which files derived from which other files. So we came up with the idea of a RAID diagram. ¶

This (Figure 7) is an extract from a RAID diagram we did for an investigation into machining cryogenically frozen materials. You can see fairly clearly where the machining parameters came from and which runs these two images came from. It's



**Figure 5:** Files and directories relating to the 2nd Year Snowmobile Design Task data case from the KIM Project

#### Project Data Record Manifest Template for IdMRC Projects

The Project Data Record Manifest (PDRM) constitutes the principal contract through which the records relating to research project may be identified and measured. It must be located in a publicly accessible and searchable place. The default location is an anonymous log in page of the research project site.

The Project Data Management Plan and the Project Data Record Manifest should be considered a pair, and should be co-located.

The PDRM should be 'read-only', writing rights being limited to members of the originating research project team and by other nominated individuals such as the data manager. A warning system must be in force.

Whilst the PDRM will be globally available, there will be some records associated with the research project which are confidential or otherwise. Access to security of this system must be limited by placing the records in appropriately password-protected locations. This could be BUCS file space or within the research project web or other space. If in doubt, the advice of the data manager (or talking back, the project PI) should be sought.

#### Summary of Research Activity

<b>Project name</b>	e.g. Long And Technical Tutorial Evaluation (LATTE)
<b>Period of Project</b>	e.g. October 2009 – March 2011
<b>Lead and partner organisations</b>	e.g. University of Bath (lead), University of Cambridge, University of Leeds
<b>Principal Investigator (name and contact details)</b>	Name: Contact details:
<b>Data access summary</b>	Data access refers to the physical means by which access to records is constrained. The overarching data access provisions for this research project are recorded in the DMP. For details of confidentiality levels of individual records see the Project Data Record List below. As a guide, data access should be either consistent with or more restrictive than the confidentiality status.
<b>Retaining repository</b>	The data from this Research Activity will be deposited according to the IdMRC DMP (see below) or The data from this research activity will be deposited in ....
<b>Related documentation</b>	<ul style="list-style-type: none"> <li>IdMRC Policy and Code of Conduct on the Governance of Good Research Conduct</li> <li>The University of Bath Good Research Practice for Research</li> <li>Engineering Research Data Management Plan Specification</li> <li>IdMRC Project Data Management Plan</li> </ul>

#### Project Management Documentation

Note that some of these records may need to be placed in a password-protected storage area.

- Project Data Record Manifest (this file)
- Project Plan (this file)
- Confidentiality agreement with (partner) (this file: note if this agreement is itself confidential it should be placed in an appropriately password-protected location)
- Confidentiality agreement with (partner) (this file: note if this agreement is itself confidential it should be placed in an appropriately password-protected location)
- Confidentiality agreement with (partner) (this file: note if this agreement is itself confidential it should be placed in an appropriately password-protected location)
- Confidentiality agreement with (partner) (this file: note if this agreement is itself confidential it should be placed in an appropriately password-protected location)
- Confidentiality agreement with (partner) (this file: note if this agreement is itself confidential it should be placed in an appropriately password-protected location)
- Confidentiality agreement with (partner) (this file: note if this agreement is itself confidential it should be placed in an appropriately password-protected location)

#### Project Data Management Documentation

- Project Data Management Plan (this file) (this will be a regional association, since the PDRM will identify the Project Data Record Manifest)
- Other data record-associative documents (this file)

#### Project Data Record List

Every project data record should be listed in the table below in the form: Title, file name, record type, location, owner and contact details, confidentiality status.

#### Record Type (for both electronic and physical records)

Every data record will be one of the following: research data record, context data record, associative data record, research object data record, experimental apparatus data record.

#### Location

If all the files are archived in a single, central location, the location need be identified for the set of records (the Data Class) only. For electronic records it is expected that a hyperlink or filepath to the location is recorded. For physical records the location should be described.

#### Owner

The owner is the person currently responsible for the management of the record, and who is in a position to consider matters such as integrity and security. Ownership does not imply any rights to use or dispose. During the period that the research project is under way it is likely that the owner will be a research officer or an individual in a temporary role. If project and the ownership should be transferred to an external individual, such as the project PI or the data manager responsible. In many cases it will be appropriate for a research officer to retain ownership.

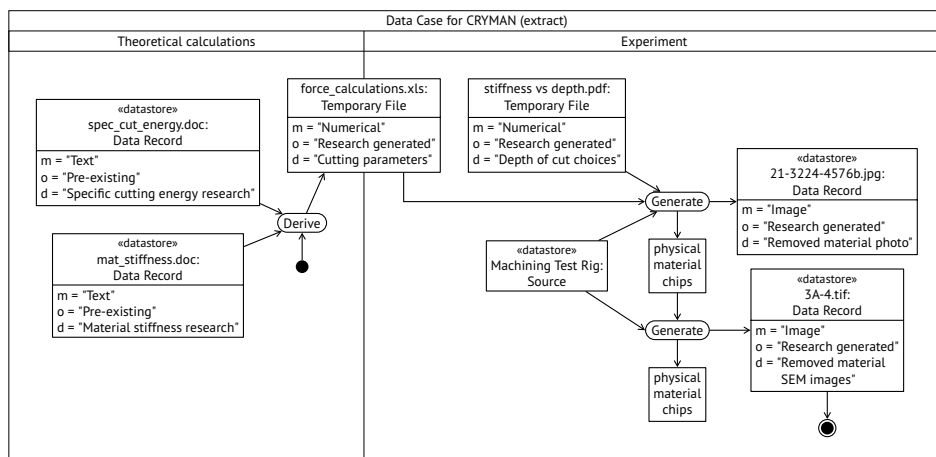
#### Confidentiality Status

Confidentiality status indicates what classes of people and what automated information-gathering systems may have sight of the data record. It does not indicate information about how the records are protected. It is likely that the confidentiality status will change during the life cycle of the data record, in which case the status must be updated. Access to the data must be restricted. If access is free, then the term 'public domain' should be used. If the access is restricted, then the access should be described either by naming groups or individuals.

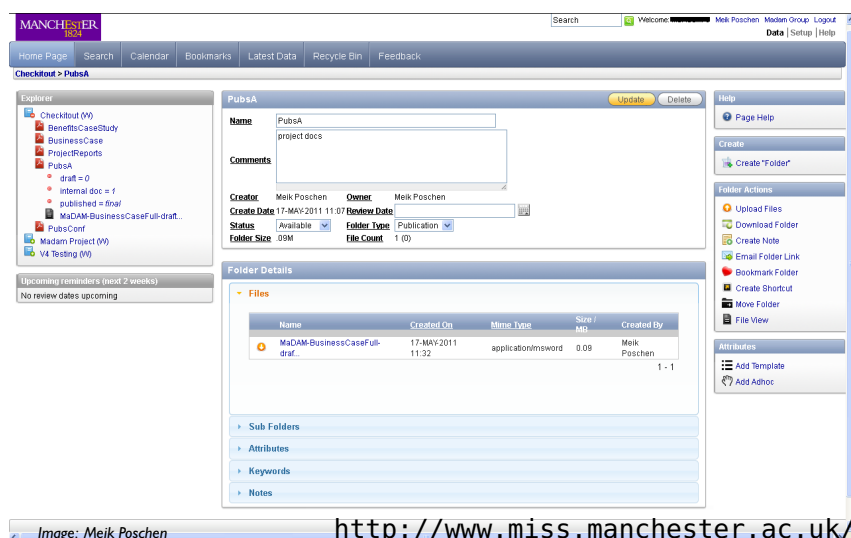
Record Title	File Name	Owner	Contact Details	Data Record Type	Confidentiality Status
Example:					
IdMRC Research Project Data Record	idmrcrpt1021.rpt	Minor	minor@bath.ac.uk	associative data record	public domain

#### History of this PDRM

**Figure 6:** Project Data Record Manifest Template for IdMRC Projects



**Figure 7:** Example RAID diagram



**Figure 8:** Web-based interface to the MaDAM system

better, but still a bit cumbersome to do by hand, so we put together a tool called RAIDmap to make it easy. ¶ It is based on the Open University's Compendium tool, with a few bells and whistles added, and while it's a bit rough around the edges it is available for anyone to download and use. ¶

Our project wasn't the only one producing useful data management tools. Jisc funded a whole programme of such projects. Among them was MaDAM at the University of Manchester (¶ Figure 8). What you can see on the slide is the web interface to a research data storage system. The top level folders relate to projects, with lower levels representing experiments, samples, publications and so on. Access controls may be set at both directory and file level, and for both groups and individuals. The system encourages researchers to add metadata to their files, and will even help them package their data up for submission to the institutional repository.

Another was DataFlow at the University of Oxford. It produced a pair of tools called DataStage and DataBank. DataStage, like MaDAM, is the working collaboration environment while DataBank is an institutional data repository.

I was also involved in a project called the Smart Research Framework, which developed tools that incorporate data management directly into researcher workflows. ¶ One of the tools is an electronic lab notebook system called LabTrove. ELNs are often used by industrial research chemists but we're also seeing them in academia and in engineering. LabTrove is based on blog technology, where each post can represent a sample, a technique, a methodological stage, or the data output from a particular run, and they all link together to make a highly efficient and easily navigable scientific record. Standalone instances of this technology have been installed in various places and it can also be used as a cloud application.

Those are just a few of the tools now available to help researchers perform research data management.<sup>14</sup> To conclude, I'd like to look at what role you as librarians might be able to play.

<sup>14</sup>More can be browsed at <http://www.dcc.ac.uk/resources/external/tools-services>

## RDM and libraries

The DCC is running a series of Institutional Engagements to support universities to develop RDM services...<sup>15</sup>

Based on the 21 universities that we are working with, we have found that the library leads RDM initiatives in the vast majority of cases. Research offices, IT teams and other services such as records management are also involved, but often to a lesser degree. University libraries have been instrumental in many areas of work:

- Making a business case for RDM to senior managers
- Defining institutional RDM strategy
- Developing institutional and departmental RDM policy
- Delivering training courses
- Helping researchers to write DMPs
- Advising on data sharing and citation
- Setting up data repositories and catalogues
- Auditing the institution's data holdings
- ...

Librarians are well-placed to support researchers:

- they have a highly relevant skill set:
  - they have knowledge of information management, metadata, etc.
  - they have experience of teaching information literacy
  - they already run publication repositories
  - they have proven liaison and negotiation skills
- they have good relationships with researchers *so it easier to provide support*
- they have existing open access leadership roles *so are often expected to address RDM too*

So how can you get up to speed with research data management?

Digital Curation Centre

- Briefing papers: <http://www.dcc.ac.uk/resources/briefing-papers>
- How-to guides (*see especially the one on developing research data management services*): <http://www.dcc.ac.uk/resources/how-guides>

---

<sup>15</sup><http://www.dcc.ac.uk/community/institutional-engagements>



## Training materials aimed at librarians

- MANTRA: <http://datalib.edina.ac.uk/mantra/libtraining.html>
- RDMRose Lite: [http://rdmrose.group.shef.ac.uk/?page\\_id=364](http://rdmrose.group.shef.ac.uk/?page_id=364)
- SupportDM: <http://www.uel.ac.uk/trad/outputs/resources/>

*Alex Ball. DCC/UKOLN, University of Bath. <http://www.ukoln.ac.uk/ukoln/staff/a.ball/>*

---



Except where otherwise stated, this work is licensed under Creative Commons Attribution 2.5 Scotland: <http://creativecommons.org/licenses/by/2.5/scotland/>



The DCC is funded by JISC.

For more information, please visit <http://www.dcc.ac.uk/>